

## 4.0 Water Monitoring: Validation and Data Quality Assessment

Data validation and verification (V&V) during CY 2006 was performed by LM personnel at the Grand Junction Office. Data quality assessment (DQA) is performed by personnel at the Site. The following section distinguishes DQA from data validation, and discusses the technical basis, equations, and criteria used for DQA of water.

### 4.1 General Discussion

Data validation is the principal means of assessing the usability of water analytical data. Validation also improves overall data quality by allowing the laboratory coordinator to closely monitor laboratory performance and to provide feedback to each laboratory regarding its ability to produce quality data that meets subcontract requirements. The laboratory coordinator may also use the results of data validation to direct analytical work to laboratories that demonstrate superior performance by generating timely, high quality analytical data for the Site.

Data validation is a rigorous data review performed by the laboratory coordinator or designee on all of the water analytical data generated by the Site. Additionally, the Site lead may request a secondary detailed validation on a case-by-case basis. Data validation is currently performed as specified in Stoller procedure *Environmental Procedures Catalog* (STO 6), "Standard Practice for Validation of Laboratory Data," GT-9(P). This procedure is based on the following EPA documents:

- U.S. Environmental Protection Agency, July 2002, USEPA Contract Laboratory Program National Functional Guidelines for Inorganic Data Review, EPA540/R-01/008.
- U.S. Environmental Protection Agency, October 1999, USEPA Contract Laboratory Program National Functional Guidelines for Organic Data Review, EPA540/R-99/008.
- U.S. Environmental Protection Agency, June 2001, USEPA Contract Laboratory Program National Functional Guidelines for Low Concentration Organic Data Review, EPA540/R-01/006.
- U.S. Environmental Protection Agency Office of Environmental Management, April 1997, Evaluation of Radiochemical Data Usability, ES/ER/MS-5.

All water analytical data collected by the Site are considered valid unless data validation identifies analytical problems that require the data to be qualified. When it is necessary to qualify individual data records, standard qualifier codes (alphanumeric validation codes) are applied. Integer "reason codes" often accompany these validation codes, enabling the data user to determine why the results were qualified.

Common data qualifiers used by LM are defined below. Refer to *Environmental Procedures Catalog* (STO 6), "Standard Practice for Validation of Laboratory Data," GT-9(P) for formal definitions.

- U The material was analyzed for but was not detected. The associated numerical value is the sample quantitation limit.
- J The associated numerical value is an estimated quantity.

- R The data are unusable (compound may or may not be present). Re-sampling and reanalysis are necessary for verification.
- N Presumptive evidence of the presence of the material.
- NJ Presumptive evidence of the presence of the material at an estimated quantity.
- UJ The material was analyzed for but was not detected. The sample quantitation limit is an estimated value.

Data validation includes the evaluation of laboratory QC data such as method blanks, laboratory control samples (LCS), and spike recoveries. Adherence to sample and extract holding times, standard analytical methods, contractual requirements, and proper documentation are also verified.

Although DQA and data validation examine some of the same QC data, they do so from different perspectives. DQA (in this report) looks at the overall quality of an entire year of water data, in contrast to validation, which looks at the analytical details of individual data packages. Data validation focuses on laboratory performance, while DQA focuses on interpretation of data describing QC samples that originated in the field, such as “field duplicate” samples and “equipment rinsate” samples.

In contrast to data validation, the DQA performed by personnel at the Site, does not assign data qualifiers to individual analytical results or data packages. DQA is a second level of QA intended to be a general assessment of how well the water data-collection program is operating. The DQA is performed by evaluating water quality data in terms of the precision, accuracy, representativeness, completeness, and comparability (PARCC) parameters.

## 4.2 PARCC Parameters

Use of the PARCC parameters for DQA has been promoted by EPA guidance documents. Accuracy and precision are quantitative measures. Representativeness and comparability are qualitative measures. Completeness is a combination of both quantitative and qualitative measures.

Site personnel evaluate the PARCC parameters by following guidelines published in the following QC documents.

- RMRS, 1998, Procedure for Evaluation of Data for Usability.
- RMRS, 2000, Quality Assurance Program Plan for the Automated Surface-Water Monitoring Program. RF/RMRS-2000-013, Revision 0, March 2000.
- RMRS, 2001, Quality Assurance Program Plan for the Groundwater Monitoring Program Rocky Flats Environmental Technology Site.

The following subsections discuss the PARCC parameters and the types of data available to assess them.

### 4.2.1 Criteria for Precision

The precision of a measurement is an expression of the mutual agreement between duplicate measurements of the same property taken under similar conditions. Precision can be expressed quantitatively by the relative percent difference (RPD) between real and field duplicate samples for metals, VOCs, polychlorinated biphenyls and water quality parameters (WQPs) as defined by the following equation:

$$RPD = \frac{|(S - D)|}{(S + D)/2} * 100$$

where: S = Concentration of analyte in Real Sample  
D = Concentration of analyte in Duplicate Sample  
Undetects are not included

The Site uses the “Duplicate Error Ratio” (DER) to quantify the precision of radionuclide activity data:

$$DER = \frac{|(S - D)|}{\sqrt{[(TPU_S)^2 + (TPU_D)^2]}}$$

where: TPU<sub>S</sub> = Total Propagated Uncertainty of the Sample  
TPU<sub>D</sub> = Total Propagated Uncertainty of the Duplicate  
S = Sample Result  
D = Duplicate (or Lab Replicate) Result

Because TPU is seldom reported with radionuclide activity data, the two-sigma error or random counting error has been substituted for TPU in the U, Am, and Pu calculations made for this report.

The Site QC criterion for water RPDs is that individual RPDs should be ≤30 percent. The analogous criterion for DERs is to be ≤1.96. The overall goal for the water data set is to have 85 percent of the RPD and DER values comply with the QC criteria.

### 4.2.2 Criteria for Accuracy

Accuracy is the degree of agreement for a measurement with an accepted reference or true value and is a measure of the bias in a system. The closer the measurement is to the true value, the more accurate the measurement. The Site validation process is the principal means for evaluating the accuracy of analytical results.

Because the Site V&V process compares the actual analytical methods used by each laboratory to the contract-required analytical methods, the Site does not repeat this evaluation.

Matrix spike (MS) and matrix spike duplicate (MSD) recoveries are reported by the analytical laboratories for most non-radionuclide analytical suites. Criteria for acceptable MS recoveries vary between laboratories, depending on the analyte, and the analytical method. The Site criterion for acceptable MS results ranges from 75 to 125 percent recovery.

LCS recoveries for radionuclides are often available for water-quality data. Laboratories in practice will commonly accept LCS values in the range of 70–130 percent. LCS percent recoveries between the 70–130 percent laboratory range and the 75–125 percent QC range required by the Site laboratory contracts are examined by data validators for acceptability on an analyte-by-analyte basis. The Site criterion for acceptable LCS recoveries ranges from 75 to 125 percent recovery.

Because some laboratories reported LCS results in pCi/L, while others calculated percent recovery, the Site uses the “relative bias” reporting criterion. The relative bias criterion is defined in the BOA by the following formula (see Page J-6 of the National BOA, Section 2.3.2.5):

$$\text{Relative Bias} = (\text{Observed} - \text{Known}) / \text{Known}$$

where: Observed = measured activity of LCS standard (pCi/L)

Known = known activity of LCS standard (pCi/L)

Acceptable values for relative bias results range from –0.25 to +0.25.

#### **4.2.3 Criteria for Representativeness**

Representativeness in DQA is limited to an evaluation of whether analytical results for field samples are truly representative of environmental concentrations, or whether they may have been influenced by the introduction of contamination during collection and handling. The potential introduction of contamination is commonly evaluated by examination of the analytical results for equipment rinsates.

Equipment rinsates are used to assess the efficacy of the decontamination process used to clean water sampling equipment. Analytes detected in rinsate samples indicate possible cross-contamination between environmental samples. Rinsates are samples of volatile-free “distilled” water that have been poured over or through decontaminated sampling equipment and subsequently handled in the same manner as environmental samples. For flow-paced composite samples that are collected over time in carboys, a location-specific “rinse carboy” is prepared using distilled water. This carboy is treated the same as other surface-water samples from that location, and analyzed for the same parameters. Analytical data from these rinse carboys are used to assess how well the carboys were cleaned between field deployments and to determine if contamination was introduced during sample preparation.

Although rinsates are used specifically as indicators of cross-contamination from improper decontamination of equipment, they are carried through the entire sampling, shipping, and laboratory process. Therefore, they are good indicators of potential contamination introduced during any of these steps.

#### **4.2.4 Criteria for Completeness**

A qualitative measure of completeness is the rate of successful sampling. The DQA verifies that all planned samples were collected, unless insufficient water was available for sampling. The completeness goal for successful sampling is the collection of at least 90 percent of the planned samples. However, the availability of water is outside the control of the Site. If all required stations were visited, sampling completeness is considered acceptable.

Completeness as a quantitative measure of data quality may be expressed as the percentage of valid or acceptable data obtained from a measurement system. The Site tracks analytical laboratory performance through both the shipment of samples to the laboratory and the receipt of data from the laboratory. The also evaluates data completeness on the following formula:

$$Completeness = DP_u = \frac{DP_t - DP_n}{DP_t} * 100$$

where:  $DP_u$  = Percentage of usable data points

$DP_t$  = Total number of data points

$DP_n$  = Non-usable (rejected) data points

The completeness criterion is having  $\geq 90$  percent valid samples.

#### 4.2.5 Criteria for Comparability

Comparability is a qualitative parameter. Consistency in the acquisition, handling, and analysis of samples is necessary for comparing results. Samples are collected in accordance with Site standard operating procedures (SOPs), transported per Site SOPs and U.S. Department of Transportation shipping regulations, and analyzed using standard EPA, or nationally recognized analytical methods. This helps to ensure comparability of results with other analyses performed in a similar manner.

The laboratory coordinator or designee verifies that laboratory analyses are performed according to the standard protocols specified by the Site subcontract to each laboratory. Therefore, the analytical results should be comparable to data produced by similar methods.

### 4.3 Water DQA Results CY 2006

During CY 2006, 119 locations were sampled one or more times. This resulted in a total of 399 water samples collected.<sup>38</sup> During CY 2006, 1,099 bottles of water were submitted to analytical laboratories for analysis. The following table breaks this data down by sample type.

*Table 4-1. CY 2006 Sample Type Breakdown*

	Unique Water Samples	Unique Bottle Codes
<b>Primary samples (REALs)</b>	383	988
<b>Field duplicates (DUPs)</b>	16	54
<b>Rinsates (RNSs)</b>	12	57
<b>Totals</b>	411	1099

Data used to evaluate the PARCC parameters are included in the available CY 2006 analytical data generated by the laboratories. These include analyses of field duplicate and rinsate QC samples submitted to the laboratory, and laboratory generated QA/QC samples such as LCS. The DQA of these analyses is discussed below by each PARCC parameter.

<sup>38</sup> This is the sum of real and duplicate samples for unique sampling events.

### 4.3.1 Precision During CY 2006

DERs are indicators of precision for radionuclide analyses. The QC criterion for precision requires that individual DER values should be  $\leq 1.96$ , and overall the data set should have  $\geq 85$  percent compliance with the criterion. Appendix Table B-1 is a tabulation of the DER values for CY 2006 radionuclide analyses. The table has been sorted by the DER parameter so that the range of values is apparent. The DER range is from 0.01 to 1.63.

Table 4–2 summarizes the DER findings of Table B-1 and indicates if the 85 percent goal has been met. Overall, 100 percent of the DER data are in compliance with the criterion, indicating excellent precision for radionuclide analyses.

*Table 4–2. Summary of Duplicate Error Ratio Values*

Analyte Group	Total Number of DER Results	Number of Unacceptable Results DER > 1.96	Number of Acceptable Results	Percentage Acceptable	Goal Met
Radionuclides	27	0	27	100%	Yes

RPD between real and field duplicate sample results is an indicator of precision for non-radionuclide analyses. Individual RPD values should be  $\leq 30$  percent and at least 85 percent of the RPDs should comply with the criterion. Appendix Table B-2 tabulates RPD values and is sorted first by analyte suite, then by RPD, in order to highlight the RPD range of each suite. RPD values ranged from 0.0 percent to 98.18 percent for metals, 0.66 percent to 2.67 percent for WQPs, and 0.0 percent to 70.79 percent for VOCs/SVOCs.

Table 4–3 summarizes the RPD findings of Table B-2 and indicates if the 85 percent goal has been met. During CY 2006, the RPD goal was met for metals and for all analyte groups. Overall, the non-radionuclide data had 93.3 percent acceptable RPDs, and therefore exceeded the 85 percent goal.

*Table 4–3. Summary of RPD Values*

Analyte Group	Total Number of RPD Results	Number of Unacceptable Results RPD > 30%	Number of Acceptable Results	Percentage Acceptable	Goal Met
Metals	8	1	9	88.9	Yes
WQP	4	0	4	100	Yes
VOC/SVOC	33	2	31	93.9	Yes
Totals	45	3	42	93.3	Yes (overall)

### 4.3.2 Accuracy During CY 2006

MS recoveries provide another measure of accuracy. Appendix Table B-3 displays recoveries for 479 MS and MSD analytical records for metals, VOCs/SVOCs, and WQPs. This data is summarized in Table 4–4. All individual suites, except for WQP, met the goal with greater than

90 percent of their spike recoveries falling in the acceptable range. Overall, across all analytical suites, the percentage of acceptable MS/MSD results was 96.5 percent.

*Table 4–4. Summary of MS and MSD Recovery Data*

Analyte Group	Total Number of MS & MSD Results	Number of Low Results Below 75%	Number of High Results Above 125%	Number Acceptable	Percentage Acceptable	Goal Met
Metals	302	0	0	302	100.0	Yes
WQP	47	12	0	35	74.5	No
VOC/SVOC	130	4	1	125	96.2	Yes
Totals	479	16	1	462	96.5	Yes (overall)

Appendix Table B-4 contains 33 relative bias values for LCSs. These are used by the Site to evaluate the accuracy of radionuclide analyses. The QC criterion for the acceptable range of relative bias values is from –0.25 to +0.25. During the CY 2006, the bias ranged from –0.123 to +0.140. All of the data met the QC criterion.

LCS results for non-radionuclide suites were available for metals, VOCs/SVOCs, and WQPs (including anions). These LCS recoveries are tabulated in Appendix Table B-5, which is sorted by analyte group, then by percent recovery. There are 260 LCS data records for metals. All of the LCS recoveries for metals fell in the range 82.4 percent to 120.4 percent and were within the 75 percent to 125 percent acceptable QC range. There are 153 LCS data records for VOCs/SVOCs. LCS recoveries for VOCs/SVOCs fell between 28 percent and 120 percent. Thirty-six records are outside the 75 percent to 125 percent acceptable QC range (76.5 percent acceptable). There are 52 LCS data records for WQPs. LCS recoveries for WQPs fell between 90 percent and 110 percent and were all acceptable. Overall for non-radionuclides, 92.3 percent of the LCS recoveries indicate that CY 2006 water analytical data for metals, VOCs/SVOCs, and WQPs are of high accuracy.

Another aspect of accuracy is “rejected data.” Out of 12,413 analytical records representing reals, duplicates and rinsates during CY 2006, eight records were rejected (R qualified) during data V&V. Another way to state this is that 99.9 percent of the analytical data collected during the year were considered to be valid and usable. Appendix Table B-6 lists the eight rejected records.

### **4.3.3 Representativeness During CY 2006**

As defined earlier, representativeness is an evaluation of the sampling procedure for its ability to reflect the true concentrations of contaminants in water. Equipment rinsate samples ( and ‘rinse carboys’) are used by the Site to determine whether there is introduced contamination from improper or incomplete decontamination of the sampling equipment.

During CY 2006 a total of 739 rinsate analytical records were generated for metals, radionuclides, VOCs/SVOCs, and WQPs. The majority of these records lack evidence of contamination. The remaining 62 records are tabulated in Appendix Table B-7. Four of these are B-qualified inorganic data which constitute only weak evidence of contamination. The “B” qualifier for inorganics indicates that the concentrations are above the instrument detection limit,

but below the method detection limit. Thirty records are J-qualified indicating an estimated quantification/result. Three of these are B-qualified organic data which constitute only weak evidence of contamination. The “B” qualifier for organics indicates that the analyte was also found in the method blank.

Only 25 records (3 percent; at the top of Table B-7) provide substantial evidence of inadequate decontamination of a sample carboy or equipment. Overall, there is very little evidence of introduced contamination during CY 2006 water sampling and/or shipping activities. Most of the 739 rinsate records appear to be clean. Therefore, water quality data for the year are judged to be representative of the actual water concentrations.

Because all required sampling locations were visited, and the samples that could be collected were analyzed, analyses for the year are judged to be representative with respect to spatial coverage.

#### 4.3.4 Completeness During CY 2006

If sufficient water is available for sampling, the goal is to have  $\geq 90$  percent successful sampling of all required locations. However, the availability of water is beyond the control of the samplers. Surface water monitoring during CY 2006 targeted sampling at 20 surface water sampling locations. In actuality, samples were collected at each of the 14 sites and were submitted to the laboratory for analysis; six locations were dry for the entire year. Ground water monitoring during CY 2005 targeted sampling at 100 wells. In actuality, samples were collected at 96 wells and were submitted to the laboratory for analysis; four locations were dry for the entire year. Treatment system monitoring during CY 2006 targeted sampling at six locations. In actuality, samples were collected at all six locations and were submitted to the laboratory for analysis. Therefore the sampling success rates for surface water, ground water, and treatment systems were 70 percent, 98 percent, and 100 percent, respectively.

V&V completeness is summarized in Table 4–5. This table compiles by analyte group, the total number of data points for reals, duplicates and rinsate samples. It then subtracts rejected data points, and subtracts points that lack validation qualifiers. The result is the net number of usable validated or verified data points, and this is expressed as percent usable data, or percent V&V completeness. The QC goal for completeness is  $\geq 90$  percent.

*Table 4–5. Summary of Validation and Verification Data Completeness*

Analyte Group	Number of Data Points	Number of Unvalidated Points	Number Rejected	Net Usable Points	Percent Completeness	Goal Met
Metals	1,163	0	0	1,163	100.0	Yes
Radionuclides	418	0	0	418	100.0	Yes
WQP	191	0	0	191	100.0	Yes
VOC/SVOC	10,641	0	8	10,633	99.9	Yes
	<b>Sum of Number of Data Points</b>	<b>Sum of Number of Unvalidated Points</b>	<b>Sum of Number Rejected</b>	<b>Sum of Net Usable Points</b>	<b>Overall Completeness</b>	<b>Goal Met</b>
<b>Totals</b>	<b>12,413</b>	<b>0</b>	<b>8</b>	<b>12,405</b>	<b>99.9</b>	<b>Yes</b>



Validation completeness for all suites exceeded 99.9 percent and exceeded the completeness goal. Therefore, from the perspective of V&V completeness, the CY 2006 water data are acceptable.

Another measure of completeness is that an adequate number of QC samples (field duplicates and equipment rinsates) must be collected to meet QC requirements. The recommended frequency for collecting duplicate samples is one duplicate (DUP) per 20 or fewer primary (REAL) water samples. In other words, duplicates should be collected at a 5 percent or greater frequency per REAL sample. Like duplicates, rinsate samples (RNS) are also to be collected at a 5 percent or greater rate.

The sample collection frequencies of REAL, DUP, and RNS samples are tabulated by analyte group in Table 4–6. The ratios of REAL/DUP samples shown in Table 4–6 meet water program QC goals with one DUP per 15.0 REALs. Across all analyte suites and samples collected during the year, the overall frequency of duplicates was 6.68 percent, exceeding program goals ( $\geq 5$  percent).

The ratios of REAL/ RNS samples in Table 4–6 also meet program QC goals with one rinsate per 14.8 REALs. Overall, across all suites and samples collected during the year, the rinsate collection frequency was 6.75 percent, exceeding program goals ( $\geq 5$ ).

*Table 4–6. Summary of Field Quality Control Samples and Data Records*

Analyte Group	Number of Locations Sampled for REALs	Number of Locations Sampled for DUPs	Ratio REALs/ DUPs (Goal <20)	Ratio REALs/ RNSs (Goal <20)	Number REAL Records	Number DUP Records	Number RNS Records	Total Records
Metals	77	9	22.5	18.6	1,059	47	57	1,163
Radionuclides	19	4	13.6	14.6	366	27	25	418
WQP	42	5	29.8	29.8	179	6	6	191
VOC/SVOC	99	9	14.3	14.3	9,339	651	651	10,641
<b>Totals</b>			<b>15.0</b>	<b>14.8</b>	<b>10,943</b>	<b>731</b>	<b>739</b>	<b>12,413</b>
<b>Percentages</b>						<b>6.68%</b>	<b>6.75%</b>	

#### 4.3.5 Comparability During CY 2006

No significant changes were made to water sampling or analytical procedures during CY 2006. Therefore, the analytical data generated during the year should be generally comparable to corresponding analyses from previous years.

End of current text